



INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE

DEPARTMENT OF ELECTRONICS, SYSTEMS, AND INFORMATICS

COMPUTING SYSTEMS ENGINEERING

MACHINE LEARNING COURSE

TEACHER: EDNA L. GUEVARA RIVERA

BIRD STRIKE FATALITY PREDICTION ON AIRPLANE CRASHES

MODEL IMPLEMENTATION

PRESENTED BY:

MARCO RICARDO CORDERO HERNÁNDEZ, 727272

CARLOS EDUARDO RODRÍGUEZ CASTRO, 727366

NOVEMBER 22ND, 2022

AUTUMN, 2022

TLAQUEPAQUE, MÉXICO

MACHINE LEARNING COURSE

Index

Index	1
Introduction	1
Model selection and motivation(s)	2
Implementation	3
Results description.....	4
Performance comparative	6
Conclusions and pending work	7
References	8

MACHINE LEARNING COURSE

Introduction

In previous steps, the selected dataset has been cleaned through multiple programming and analysis techniques in order for it to be ready for machine learning algorithms, this with the purpose of training an effective and efficient set of models, from which one will be determined to be the best amongst them, at least for this problem.

As it has been used in the past, Jupyter [1] will act as the container/holder for the computational operations results and outcomes from the algorithms.

MACHINE LEARNING COURSE

Model selection and motivation(s)

For the analysis and comparison between results, and for class material comprehension purposes, the following Machine Learning Models will be implemented:

1. *Linear Regression (Normal)*

As it is the most common type of technique and usually one of the first concepts used to teach about ML, this widely used model will function as the main comparison and example for further upgrading in next model implementations.

Although the concepts involved in LR are fairly basic, these tools are still very useful and serve as a comparison entry point.

2. *Neural Network*

Another broadly known technique when discussing about Machine/Deep Learning. This model has gained plenty of attention over recent years, as it's being used among a great range of modern-day problems, such as facial recognition, stock market predictions, signature verification, etc. [2] Thus, making it a great opportunity for a demonstration of this model for yet another contemporary problem.

3. *Decision Tree*

Lastly, this model will be visited as an alternative to classic statistic methods, as DT's support nonlinear data and makes for a great visual resource that involves several categories/features found in the dataset of analysis. This highly customizable model allows for fine-grain knobbing/adjusting for better result outcome and can be easily compared against other models.

MACHINE LEARNING COURSE

Implementation

Model implementation can be found inside the jupyter notebook created specifically for this part of the project, giving an extensive explanation of its steps and result retrieving.

This file can be found along the current report delivery.

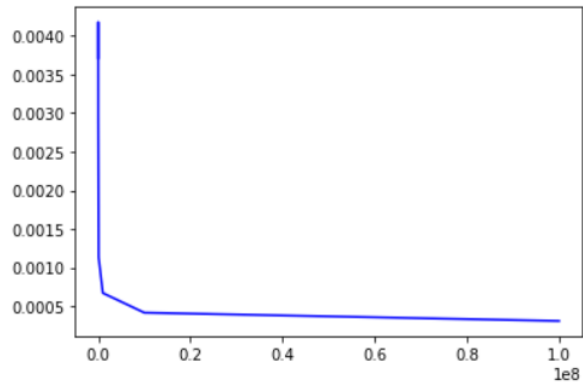


ML project_Part
3_ARPA.ipynb

MACHINE LEARNING COURSE

Results description

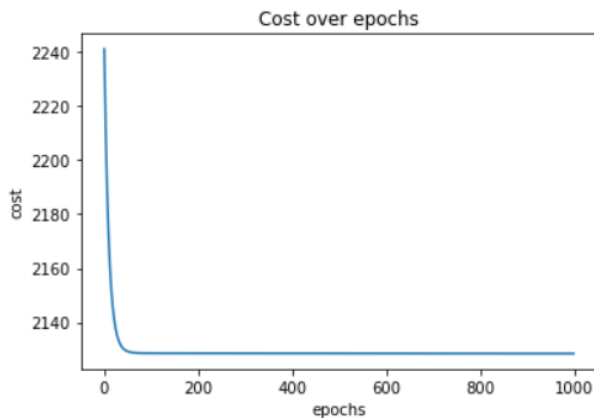
Graph generated for Linear Regression



Error (R2): 0.003701729051915792

For this model, great results have been achieved, as error is fairly low, and the graph demonstrates how cost descends elegantly, approaching zero.

Graph generated for Neural Network

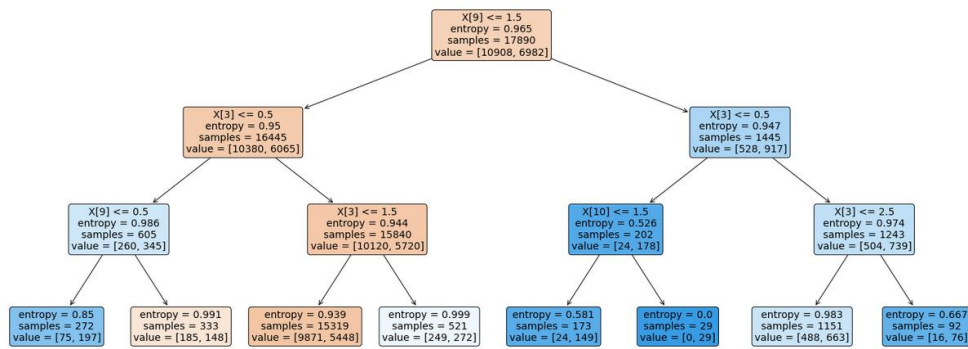


2 Neurons, accuracy = 0.6083724569640062
Initial cost: 958.7690088398112
Final cost: 913.4016468898567
3 Neurons, accuracy = 0.6083724569640062
Initial cost: 960.0093254474551
Final cost: 913.4198081944725
4 Neurons, accuracy = 0.6083724569640062
Initial cost: 959.3550561132524
Final cost: 913.424485891052
5 Neurons, accuracy = 0.6083724569640062
Initial cost: 957.1368241622101
Final cost: 913.3861474025143
6 Neurons, accuracy = 0.6083724569640062

Yet again, costs descend as epochs augment, but this is usual behavior for NN. When analyzing cost, a fluctuation can be spotted, with similar costs repeating themselves, but with zero to no accuracy upgrades. This doesn't mean something has gone wrong or similar, but perhaps this could indicate that this model may not be adequate for this particular problem.

MACHINE LEARNING COURSE

Tree graph and related data



Test matrices
[[4309 356]
[2455 548]]

[[4309 356]
[2455 548]]

Lastly, another common behavior can be spotted with this model execution when matrix analysis is made. Little under half the data it's being correctly categorized, with about 60% of accuracy. It cannot be said that this model it's bad or wrong, again, it's just not the perfect fit for what it tried to address.

MACHINE LEARNING COURSE

Performance comparative

Talking about time, the classifications remains as it follows (in ascending time order):

1. Linear Regression
2. Decision Tree
3. Neural Network

It has to be kept in mind that NN works closely with epoch concept and implementation, giving it the unfortunate last place, at least for performance.

Now, comparing results, it's easy to place LR as the best model for the current problem, as it showed tiny error, meaning virtually no cost whatsoever, but something else can be said that would be more appropriate: further analysis has to be made. Whether is dataset refining, or data manipulation, or model selection, the results gathered in this part of the project can't be totally seen as conclusive. But maybe that is what all of this it's all about, about searching and building better and better models. The results obtained are not wrong, but maybe they would be serving a greater purpose as a simple entry point.

As for now, LR it's the undeniable outstanding model. This'll be discussed in the final document.

MACHINE LEARNING COURSE

Conclusions and pending work

It can be considered now that the project has concluded, as the models for predictions has been implemented and its development has been demonstrated.

Although the conclusions along these project parts had been somewhat short and vague, a final document has to be constructed, which'll contain a more immersive conclusion, including obstacles found in the realization of this work.

With this being said, there's nothing left to do for now, as a complete and detailed deliverable it's still on its way.

References

- [1] F. Pérez y B. Granger, «Jupyter project,» Jupyter project, 2022. [Online]. Available: <https://jupyter.org/>. [Last access: 14 November 2022].
- [2] V. Kaushik, «Analytics Steps,» Analytics Steps Infomedia LLP, 26 August 2021. [Online]. Available: <https://www.analyticssteps.com/blogs/8-applications-neural-networks>. [Last access: 15 November 2022].